

長期間報道されるニュース記事を対象とした 話題推移の可視化による内容理解の支援

齋藤 藍^{a)} 松下 光範^{b)}

概要: 本研究の目的は、長期間報道されるニュースの内容理解を支援することである。近年、ニュースアプリなどの普及によりニュース記事の閲覧が容易になった反面、関連する複数のニュース記事に関連づけて理解することが困難な状況になっている。特に、長期間にわたり報道されるニュースは局面の経時的変化が起こるため、報道の初期段階から注意してニュースを閲覧していないと、そのニュースの推移を遡って理解することは容易ではない。そこで本研究では、ニュース記事の内容とその報道時期に着目し、それらを話題ごとにクラスタリングし、話題ごとの記事数を時系列で可視化することで、一連のニュースの推移を理解する支援を行う。本稿では、一連のニュースの経時的に変化する報道内容を確認できるようにした可視化インタフェースを提案する。

1. はじめに

かつて、ニュースの閲覧手段は紙媒体である新聞が中心であったが、PC やスマートフォンの普及により、現在では電子化されたニュース記事をブラウザやニュースアプリといった多様なプラットフォームで閲覧できるようになってきている。それに伴い、新聞では朝夕に限られていたニュースの配信タイミングや紙面の制限によるニュース記事数の制約が緩和されている。これに加え、電子化されたことによって過去に配信されたニュースが新聞社や配信社のサーバで保持され検索できるようになっており、その結果としてユーザがアクセス可能なニュースの量が膨大になっている。大量のニュース記事にアクセスできることはユーザにとってはメリットである反面、その膨大さゆえに情報過多に陥り、ユーザが関心を持つニュースに関する情報を効率よく収集し、全体を俯瞰しながら整理することが困難になってきている。

ニュースは必ずしも単発の記事として報道されるとは限らず、続報が報じられたり、複数の観点で異なる紙面に掲載されたりする。例えば、災害に関するニュースでは、災害発生時の第一報のあと、被害情報が明らかになるに伴って続報が報じられる。被害の程度が大きければ、続報が報じられる期間も長くなり、初期段階では被害状況に関する

内容が中心であったニュースが、時間が経つにつれ復興方針や防災対策が論じられるようになるなどその内容も変化していく。このように、長期間報道されるニュースではそのトピックが時間経過に伴って変化するため、報道の初期段階から注意してニュースを閲覧していなければ、一連のニュースの全貌を理解することが困難になる。そのため、このようなニュースを包括的に理解するには、関連する一連のニュース記事をトピックに分類し、個々のニュースをトピックごとに関連付けながら、時間の経過に伴ってそれらがどのように変化するかを把握する必要がある。

こうした観点の下、本研究では長期間報道されるニュースをトピックごとに分類し分かりやすく提示することでユーザのニュース理解の支援を目指す。その端緒として、本稿ではニュース記事の内容とその報道時期に着目し、経時的に変化する報道内容を可視化するインタフェースを提案する。

2. 関連研究

ニュース記事の理解を目的とした研究は、自然言語処理や情報可視化などの研究分野で様々な視点から行われている。

小川ら [6] は、ニュース記事の効率的な理解のために、ニュース記事の要約を行った。個々のニュース記事を3行に要約することで、ニュース記事の全文を読まずとも個々のニュース記事の概要を把握することを可能にした。この手法は、個々のニュース記事について効率的に理解できる一方で、時間経過に応じて続報や新たな問題が現れるよう

¹ 関西大学
Kansai University, 2-1-1 Ryozenji, Takatsuki, Osaka 569-1095, Japan

^{a)} k882494@kansai-u.ac.jp

^{b)} mat@res.kutc.kansai-u.ac.jp

な一連のニュースを包括的に把握するには十分とは言い難い。

鈴木ら [9] は、経済や政治などの分野で長期に渡って報道されるニュースでは、記事に登場する人物や組織が話題の推移に伴って変化する点が内容の理解を妨げていると指摘し、ニュース記事に掲載されている人物名や組織名を相関図として可視化することで、ニュースの内容理解の促進を試みている。この手法は、あるトピックに関わる個人や団体を把握することは容易になるものの、トピックの時間的な推移を包括的に理解するには十分でない。

松本ら [8] は、長期間報道されるニュース記事の理解を深めるために、ニュース記事におけるトピックの推移の抽出を試みている。この手法では、潜在的ディリクレ配分法 (Latent Dirichlet Allocation, 以下 LDA と記す) を用いて一連のニュースから期間ごとにトピックを抽出し、トピック間の類似度により各期間内のトピックを統合した。更に、連続する期間のトピック間の類似度を算出しトピックが相互に関連しているかを示した。これにより、従来の手法では観測できなかったトピックの発生と消滅を含んだトピック推移の抽出を可能にした。また、その様子をネットワーク図とワードクラウドを用いて可視化している。この手法は、経時的に変化するトピックの推移を理解することに貢献するものの、期間ごとにトピックを抽出しているため、その期間の局所的な特徴に引きずられて全体のトピック推移が分かりにくくなる懸念がある。また、LDA はひとつの記事に複数のトピックが含まれている可能性を考慮する手法であるため、記事とトピックが1対1に対応しておらず、特定のトピックについて内容を確認する際に読むべき記事数が増加する懸念がある。

谷口ら [7] は、ニュースを効率的に理解するために、ニュースのストーリーを提示するインタフェースを提案した。この手法では、ユーザの入力クエリに関する「誰が何をした」など一連のニュースにおける出来事を短文で列挙し、それを各ニュース文と共に提示することにより、一連のニュース記事の概要を把握しつつ内容を理解できるようにしている。このインタフェースでは、一連のニュースにおける出来事を網羅的に把握することが可能であるが、一連のニュースに含まれる各々のトピックの重要性や時間的な関係性など個々のニュース記事の位置づけを理解することは難しい。

これらの研究は主として新しいニュース記事が逐次的に追加される状況での利用を想定しており、すでに長期間に渡って報じられたニュースの内容を包括的に理解することを企図したものではない。このような目的の場合、長期間報道されるニュースの全体像を把握することと同時に、ユーザの興味に従って詳細な情報を確認できることが求められる。そのため、本研究では長期間報道されるニュースの概要と詳細を必要に応じて探索・理解できるようなイン

タフェースの実現を試みる。

3. デザイン指針

3.1 ニュース閲覧モデル

過去に長期間報道されたニュース記事を閲覧する際、知りたいニュースに関するキーワードで検索を行うことが一般的である。例えば、Covid-19 が与えた社会的影響をまとめるために検索を行った場合、最新のニュース記事 (e.g., θ 変異株の話題) と報道初期のニュース記事 (e.g., α 変異株の話題) が検索結果に混在し、注意深く読み解かなければ間違った理解をする懸念がある。また、対象期間を指定して検索する手法も存在するが、検索結果に出現したニュース記事は視点や論点が異なるため、自身が知りたい内容に合致した記事を効率よく見つけることは簡単ではない。

そのため、長期間報道されるニュースを効率的かつ効果的に閲覧するには、トピックごとに時系列で閲覧できるようにすることが求められる。こうした点を鑑みて、一連のニュースの内容理解を深めるための要件を以下の3点に集約した。

- ニューストピックの時間経過に伴う変化が把握できること
- ニュース全体のトピック把握と個々のニュース記事の閲覧がシームレスにできること
- 任意のトピック及び任意の期間のニュースを閲覧可能であること

本稿では、これらのデザイン指針を満たすシステムの提案と検証を行う。

3.2 情報可視化参照モデルに基づくシステムのデザイン

前節で述べた要件を Card らが整理した情報可視化参照モデル [1] (図1参照) に照らしてシステムのデザイン指針を考察する。Card らのモデルは、(1) 大量のデータ (Raw Data) から可視化に必要なデータを抽出・集約して Data Table を作成する Data Transformation プロセス、(2) 前段階の処理で得られた Data Table について、属性と可視化パラメータ (e.g., 軸, 色, 形状) とを対応付けて Visual Structure を決定する Visual Mapping プロセス、(3) 前段階の処理で得られた Visual Structure にどこに注目すべきかを反映し View を生成する View Transformation の3つのプロセスによって可視化を生成し、(4) ユーザは企図した視覚表現を得るために各々のプロセスに働きかけることでこの可視化結果に関与する、というモデルである。

以下、各々のプロセスについて、一つずつ検討していく。

3.3 想定する可視化表現

(1) Data Transformation プロセス

ニュース記事には、それらのニュース記事以外にもコラムや料理をはじめとした趣味に関する記事や季節ご

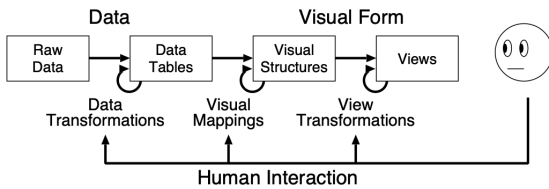


図 1 情報可視化の参照モデル [1]

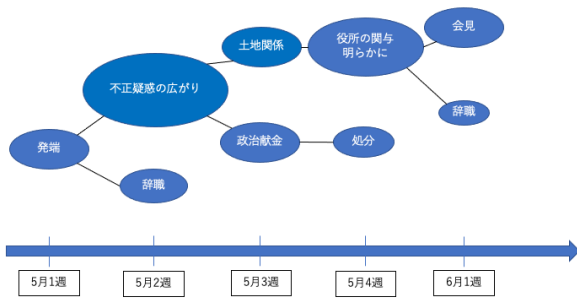


図 2 本研究が目指す可視化表現の例

との単発の記事なども含まれるため、このプロセスでは長期間にわたり報道されたニュース記事のみを、リソースとなるニュースコーパスから抽出する。抽出されたニュース記事群をトピックと期間に着目して、分類し、Data Tableに格納する。このとき、ニュース記事群がいくつかのトピックに分けるべきか事前に決定することができないため、分割数についてはデータに含まれるトピックの分布特徴からシステムが判断することとする。この手法については次章で述べる。

(2) Visual Mapping プロセス

長期間報道されるニュースは中心となるトピックが経時的に変化するため、提案システムの可視化表現ではユーザがトピックの時間的変化の概略を簡単に読み取れる必要がある。そこで、提案システムでは、一連のニュース記事を図 2 のように単位期間ごと (e.g., 1 週間, 1 ヶ月) に同一トピックの記事をノードにまとめて表現し、時系列に沿って関係するトピックを繋げて話題の分岐や収束を視覚化することで、話題推移を直観的に理解可能にする。また、報道された記事の多さは当時どれほど盛んに報道されていたかを示す指標であるため、各話題・各期間のノードに内包される記事数に従ってノードの大きさを変化させる。これにより一連のニュースにおける各トピックの注目度を把握可能にする。

(3) View Transformation プロセス

このプロセスでは、(1) 概略の理解を助けるための付加情報の表示と、(2) ユーザが関心を持った記事の詳細を確認するための個別記事の表示を行う。(1) について

は、各トピックの内容に基づいて各ノードにラベルを付与することで、一連のニュースの中で該当する時期にどのような話題のニュースが報道されたか詳細な記事を読まずに把握することを容易にする。(2) については、インタフェース画面を上下に二分割し、画面上部に可視化表現を、画面下部にニュース記事を各々表示する overview + detail [1] 方式を採用する。画面上部の可視化表現のノードをクリックすることで、そのノードに内包されているニュース記事を画面下部に表示する。このとき、画面下部のニュース記事を表示する部分には、ニュース記事の日付とタイトル、ニュース本文を表示する。可視化表現とニュース記事本文を連携させることで、ニュースの全体像の把握と詳細な内容の把握をシームレスに行うことが期待される。

3.4 想定するインタラクション

次に、このシステムを用いるユーザのインタラクションについて考える。Shneiderman は情報可視化システムを用いてユーザが情報を探索する行為のガイドラインとして Visual Information-Seeking Mantra を提案している [5]。このモデルによれば、人の情報探索行為は、まず全体を眺め (Overview first), データを拡大したり間引いたりしながら調べていき (zoom & filter), 必要に応じて詳細を確認する (details on demand), という一連のプロセスに整理できる。これに基づき、可視化システムを用いてニュースを理解する際のユーザ行為を以下のように整理した。

- (1) Overview first
 可視化表現を見ることで、一連のニュースに含まれているトピックやその時系列変化を把握する
- (2) Zoom & filter
 ノードをクリックしたり、特定の期間に焦点を当てて眺めたりすることで、トピックや時系列から閲覧する記事を絞り込む
- (3) Detail on Demand
 興味を持ったトピックから各記事内容を閲覧することで、ニュースをより詳細に理解する

4. ニュース記事の分類手法

4.1 対象とする新聞記事データ

本稿ではシステムのリソースとして毎日新聞社が提供する 2016 年版の新聞記事コーパスを使用する。この新聞記事コーパスには 2016 年に毎日新聞に掲載された記事 83,668 件が紙面での区分 (13 分野) に基づいて分類され収録されている。各々の分野の記事には特徴があり、各面によって内容が大きく異なる記事が掲載されている。例えば、家庭面 (13 面) に掲載された記事は家族の在り方や、主婦が参

考にできる料理のレシピ情報などが多く、文化面（14面）に掲載された記事は展覧会や演劇イベントを紹介する社告や新聞購読者から投稿された俳句が多い。本研究では、特定の話題が長期間継続して報道されるなかでどのようにトピックが変遷するかを観察できるようにすることを目的としているため、上述したような one-shot の記事である場合が多い家庭面や文化面の記事は対象にはなりにくい。また、スポーツ面（35面）では、野球のペナントレースの経過など長期的に報道されるニュース記事として捉えられる内容も含まれてはいるが、そこに掲載される記事は競技の結果や球団の順位、選手の成績などが中心で、時間経過に伴って異なる観点のトピックが発生することは考えにくい。そのため、本研究の対象からは除外する。

一方、毎日新聞紙面における分類のうち、経時的な局面の変化が観察できるニュース記事は国際面（7面）・経済面（8面）・科学面（16面）・社会面（41面）に多く掲載されている。国際面には、日本と外国間の外交関係や海外で起こった事件についての記事が多く、経済面には環境問題や金銭の関わる政策や市場の動向についての記事が多い。科学面には教育に関する政策や科学技術についてのニュースが多く、社会面には事件・事故や災害情報に関するニュースが掲載される。そこで、本稿ではこれら4分野（国際・経済・科学・社会）のニュースのみを対象とする。

4.2 新聞記事データの事前処理

新聞では、新聞社が注目度が高いニュースだと判断した記事は1~3面に掲載され、それ以外の記事は内容に基づいて国際面や経済面など分野ごとに掲載される。このとき、1~3面に掲載された記事は分類が明示されていないため、本稿では新聞記事紙面の分類（毎日新聞の場合は13分野）を正解ラベルとし、Support Vector Machine（以下、SVMと記す）を用いてニュース記事を学習することで分類モデルを作成し、そのモデルを用いて1~3面に掲載されたニュース記事が13分野のうちどの分野に該当するかを判断する。

分類モデル作成に用いたSVMのカーネル関数にはrbfを採用し、コストパラメータ C は1、 γ パラメータはscaleとした。SVMの実装は、python (Ver.3.8.5)のライブラリであるscikit-learn (Ver.0.23.2)*1のSVMモジュールを用いた。

新聞記事をSVMを用いて分類するための前処理として、新聞記事コーパス中の新聞記事本文を形態素に分割し名詞のみを抽出した。形態素の分割には、MeCab (Ver.0.996.2)*2およびmecab-ipadic-NEologd*3を用いた。

Slothlib[10]*4に記載されている名詞に加えて、mecab-ipadic-NEologdで未知語と設定されている名詞と1文字の名詞をストップワードに設定した。

すでに分野ごとに分類されている1~3面以外の13分野（国際・経済・社会・解説・社説・特集・総合・家庭・文化・読書・科学・芸能・スポーツ）の新聞記事の名詞にword2vec[3]を適用することで単語の分散表現を取得した。word2vecの実装においては、pythonのライブラリであるgensim (Ver.4.0.1)*5を用いた。パラメータは、ベクトルの次元数を100、windowサイズを5、反復学習回数を100、workersを3、min_countを5とした。これにより取得した単語分散表現を平均化[4]*6し、各記事をベクトル化した。1~3面以外の新聞記事には新聞の紙面情報から13分野それぞれに分割され、ラベルが付与されている。このラベルを分類基準として用い、SVMにより記事ベクトルから新聞記事の記事分類モデルを作成した。この記事分類モデルの妥当性を検証するため、1~3面以外の13分野の新聞記事を学習データと評価データに分割して5分割交差検証を行った結果の平均は、正解率は0.80、F値は0.79であった。

先述の結果から、新聞記事の記事分類モデルの妥当性が検証されたため、このモデルを用いて、内容に則した分野の分類がなされていない1~3面の新聞記事の分類を行った。新聞記事のうち1~3面に掲載された記事の単語分散表現をベクトル化し、新聞記事の記事分類モデルを用いて1~3面以外の13分野のうち該当分野に分類を行った。ここで国際・経済・科学・社会の4分野に分類された1~3面の記事は、各々1,737件、1,377件、66件、2,627件であった。本稿では、この国際・経済・科学・社会の4分野に分類された1~3面の記事と、1~3面以外の国際・経済・科学・社会の4分野の記事（以下、記事データと記す）を対象としてシステムを作成する。

4.3 トピックによる分類

本研究では、長期間報道される一連のニュースの話題の流れを可視化することで、ニュースについての理解促進を目指している。しかし、記事データはどの記事がどのような事柄について報道されているうちのひとつであるのかは判別できない。そのため、本稿ではニュース本文に任意のキーワードが含まれるデータの集合を一連のニュースと見なす。

本研究の目的である話題の分岐・収束の可視化のためには、一連のニュース記事のうちどのような話題の記事が存在するかを知ることが必要である。そのため、一連のニュースに含まれる記事を話題ごとに分類した。一

*1 <https://scikit-learn.org/stable/> (2021/12/9 確認)

*2 <https://taku910.github.io/mecab/> (2021/12/9 確認)

*3 <https://github.com/neologd/mecab-ipadic-neologd> (2021/12/9 確認)

*4 <https://ja.osdn.net/projects/slothlib/> (2021/12/9 確認)

*5 <https://radimrehurek.com/gensim/> (2021/12/9 確認)

*6 <https://github.com/yagays/swem> (2021/12/9 確認)

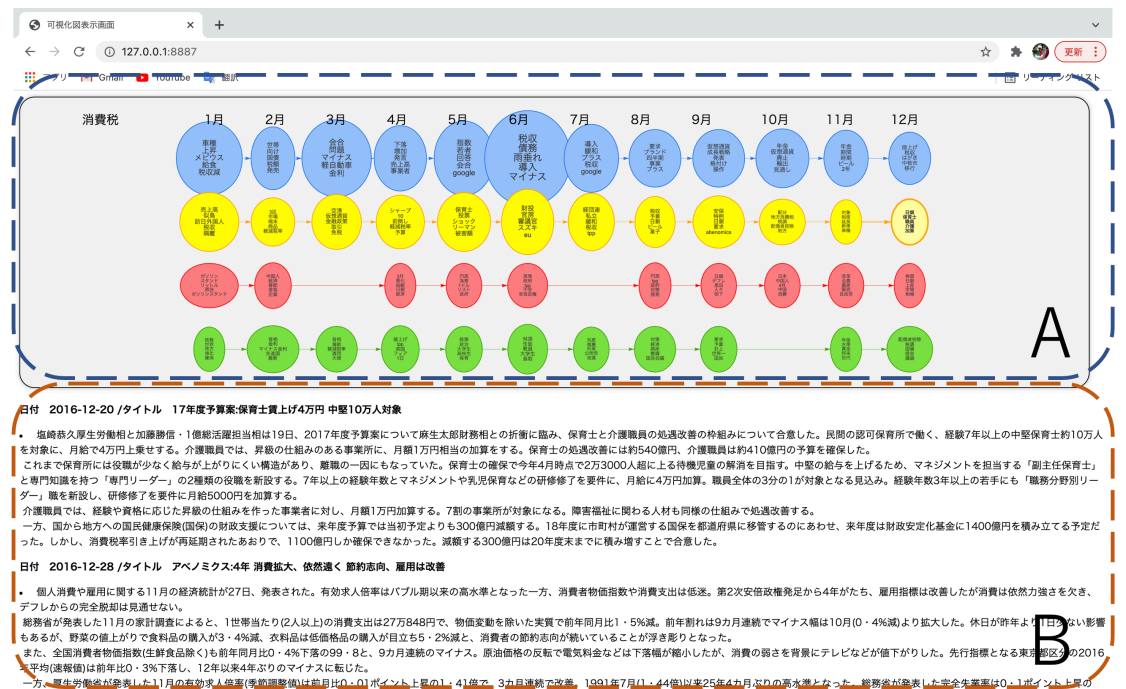


図 3 提案システムの概観。A はトピックの推移を示す可視化表現であり、ノードをクリックすることでそれに対応する記事を B で閲覧できる

連のニュースとして抽出できた記事の新聞記事本文を形態素に分割し、名詞のみを抽出した。形態素の分割には、MeCab (Ver.0.996.2) および mecab-ipadic-NEologd を用いた。Slothlib に記載されている名詞に加えて、mecab-ipadic-NEologd で未知語と設定されている名詞と 1 文字の名詞をストップワードに設定した。一連のニュースとして抽出できた記事ごとに Doc2Vec[2] を適用することで記事ごとの分散表現を取得した。Doc2Vec の実装には、python のライブラリである gensim (Ver.4.0.1) を用いた。

記事ごとの分散表現をもとにクラスタリングを行うことで、一連のニュース記事を話題ごとに分類した。クラスタリングには変分混合ガウス分布 (Variational Bayesian Gaussian Mixture, 以下、VBGMM と記す) を用いた。VBGMM の実装には python のライブラリである scikit-learn を用いた。VBGMM は、クラスタリングにおけるデータ内のクラスタ数が不明である場合に利用できるクラスタリング手法であり、一連のニュース記事データを分類する適切なクラスタ数が不明であるため、VBGMM を用いて記事ごとの分散表現からクラスタリングを行うこととした。

4.4 ノードに付与する特徴語の選択

次に、トピックの遷移を表す可視化表現のノードに表示する特徴語の選択を行う。可視化表現として採用したネットワーク図では、各トピックを月毎に表示することでトピックの遷移を観察できるようにする。そのため、各ノードには、ノード ID・ノードに該当するトピック及び月の記事中の特徴語・ノードに該当するトピック及び月の記事

数・ノードを表示する際の座標情報を付与している。ノードに該当するトピック及び月の記事中の特徴語については、各ノードに該当するトピック及び月の記事を形態素に分割し、名詞のみを抽出した上で各単語の TF-IDF 値を算出した。形態素の分割には、MeCab (Ver.0.996.2) および mecab-ipadic-NEologd を用い、TF-IDF 値の算出においては、python のライブラリである gensim (Ver.4.0.1) を用いた。TF-IDF 値が上位 5 つの単語を各ノードにおける特徴語とした。

5. 実装

5.1 システムの概要

提案システムは vis.js (Ver.4.21.0)*7 を用いた Web アプリケーションとして実装した。

実装したシステムのインターフェースを図 3 に示す。このシステムは、それぞれのトピックの経時的変化を表示する領域 (図 3-A) と、4.3 節で述べた特定の単語が含まれているニュース記事の日付とタイトル及び本文を表示する領域 (図 3-B) から構成される。本稿では、「甘利」・「参院選」・「消費税+消費増税」それぞれのキーワードで抽出した記事群を対象として説明する。

5.2 可視化表現の表示画面

図 3-A には、可視化表現を表示している。4 章で述べたトピックごとの分類を色別に表示し、横軸は時系列を表している。各ノードの大小により任意の期間中の各トピック

*7 <https://visjs.org/> (2021/12/05 確認)

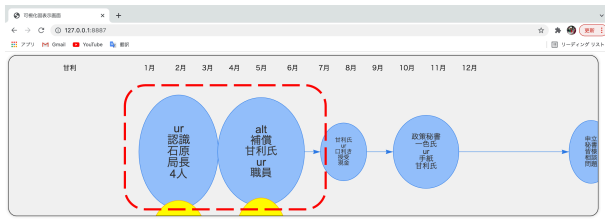


図 4 注目したトピックへのズームイン

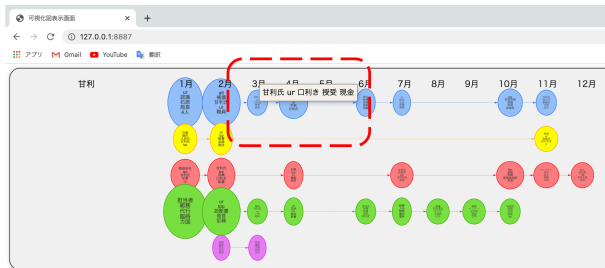


図 5 特徴語の拡大

におけるニュース記事の数を表し、ノード内には 4.4 節で作成した特徴語のうち上位 5 つを表示している。可視化表現は拡大することが可能 (図 4 参照) であり、ノード上にマウスオーバーすることで特徴語の拡大表示が可能である (図 5 参照)。これらの特徴語によって図 6 のようなトピック変遷の解釈が可能である。さらに、可視化表現の各ノードをクリックすることで、該当する期間かつトピックの記事が画面下部に表示される。

5.3 ニュース記事の表示画面

図 3-B は、ニュース記事の日付・タイトル・本文を表示している領域である。ニュース記事を可視化表現の下部に表示することで、シームレスに可視化表現とニュース記事を閲覧することができ、ニュースの詳細な理解につながることを期待される。

このインターフェースは、3 章で述べたデザイン指針に基づいて設計している。図 3-A の可視化表現によってニューストピックの時間経過に伴う変化が把握可能であり、システム上の可視化表現内のノードをクリックすることで、各ノードに対応する任意のトピック及び任意の期間のニュースを図 3-B で閲覧可能である。これらにより、ニュース全体のトピック把握と個々のニュース記事の閲覧が一画面内でシームレスに行えるようになった。

6. 実験

本実験では、ユーザが 3 章で想定したインタラクションに沿った利用をするのかについて検証する。実験では、「甘利」・「消費税 or 消費増税」の 2 つをキーワードとし、記事データより該当する一連の記事を抽出したものを実験データとした。

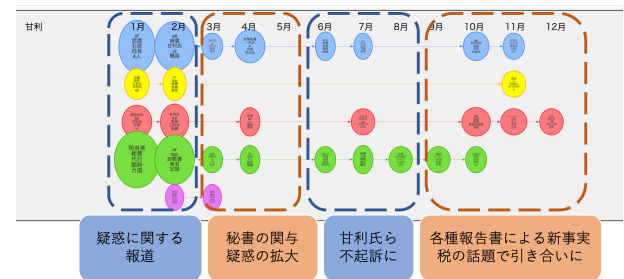


図 6 トピックの変遷の解釈

6.1 実験デザイン

本実験では、ユーザにインターフェースを使用してもらい、使用時の探索行動の観察や参考にした情報の聞き取りを通して検証を行う。ここで、実験の目的が実験参加者に伝わり、実験結果に影響が出る可能性があるため、実験参加者にはインターフェースを利用しなければならないようなタスクを設ける。タスクを設けることで、目的を持ったインターフェースでの情報探索にも繋がる。そこで今回は、インターフェースで提示される情報のみを用いてレポートを書くというタスクを設けた。実験にあたっては、各実験参加者の文章構成能力や文章作成能力の影響による過度なタスクへの集中 (情報探索の時間よりもレポート記入時間のほうが長い) が無いように、インターフェースからのニュース記事のコピー&ペーストのみでレポートを記入するように制限を設けた。さらに、コピー&ペーストのみでレポートを記入することで文脈を繋げることが困難になる可能性があるため、文脈の考慮はせずに記述することを許可した。

6.2 実験手続き

本実験では、タスクの特性上、レポートを作成する機会の多い大学生及び大学院生を実験の対象とした。本実験の実験参加者は 8 人で、キーワード「甘利」と「消費税+消費増税」のタスクに 4 名ずつ無作為に振り分けた。

まず、実験協力者に対して「実験の流れ」と「実験内タスク」について説明した。実験内タスクについては、キーワード「甘利」で抽出したデータを用いる実験参加者には、『2016 年に報じられた政治家甘利氏の疑惑について、A4 サイズ 1 ページ程度にまとめてください』という指示を行った。キーワード「消費税+消費増税」のデータを使用する実験回については、『2016 年の消費増税に関するニュースを A4 サイズ 1 ページ程度にまとめてください』とした。実験参加者には、タスク開始の前にシステムで得られる情報が 2016 年のニュース記事のうちそのキーワードを含む記事であることのみを伝えた。

ユーザが初見の状況下で可視化表現から得た情報を調査するために、システム使用前には可視化表現に関する説明を行わず、システム使用開始の 5 分後に説明を行った。可視化表現に関する説明では、(1) 横軸が時間軸であること、

(2) トピックごとに色分けがされていること, (3) 特徴語がそのノードを代表する語であること, (4) ノードの大きさが記事数の多寡を示すことについて説明を行った. その後, 時間制限は設けず, タスクが完了するまでシステムを使用してもらった. タスク完了後に (1) 普段のニュース視聴の程度, (2) , (3) インタフェースの使用感について聞き取り調査を行った.

6.3 実験結果

キーワード「甘利」のレポートについて, 完了までにかかった時間は平均 26.1 分 (max 35.7 分, min 18.5 分) であった. キーワード「甘利」のレポートでは, 4 人中 4 人が疑惑の経過について記述していたが, 結果として, 甘利氏本人及び秘書が共に不起訴となったという言及に至っているのは, 4 人中 2 人だった. また, 4 人中 3 人が甘利氏の経済再生担当の辞任について記述していた.

キーワード「消費税+消費増税」のレポートについて, 完了までにかかった時間は平均 27.0 分 (max 30.4 分, min 23.3 分) かかった. キーワード「消費税+消費増税」のレポートでは, 4 人中 3 人が税率の引き上げ及び消費増税の延期について記述していた. 軽減税率について言及したのは, 4 人中 2 人だった.

探索時にノードをクリックする行為に着目し, 実験参加者の探索の傾向を調べた. 実験参加者 8 人全員のノードクリックのうち, 平均 43.7% がノードクリックの前に閲覧していたニュース記事よりも新しいニュース記事が閲覧できるノードをクリックしていた. これは時系列順に内容を確認していることを示唆している. また, 平均で 24.8% のクリックが同時期の異なるトピックを表すノードをクリックしていた. これは同時期にどのようなトピックが報道されていたかを比較していることを示唆している.

ノードの大きさが探索行為に与える影響を調べるために, それぞれのノード内記事数をもとに大きいノードのクリック数を調べた. それぞれのデータごとに含まれるニュース記事のうち 10% が含まれるノードについて, ノードが大きいという判断をした. キーワード「甘利」は全 29 ノードのうち 4 ノードが, キーワード「消費税+消費増税」は全 45 ノードのうち 1 ノードが該当した. 実験参加者の全クリックのうち, 13.9% が該当するこれらの大きいノードをクリックしていた. 全ノードのうち大きなノードが占める割合は 6.8% であり, 大きなノードの個数に比べて大きなノードを探索した回数の割合が高いことから, ノードの大きさが探索に与える効果が確認された.

聞き取り調査の回答については, (1) 普段のニュース視聴の程度については, 8 人中 6 人が 1 日平均 1 時間未満だった. このことから, 実験参加者のニュースに関する事前知識が乏しく, 今回のニュースに関しても知識がほとんどない状態で実験参加していたことがわかる. (2) システムの

可視化表現の理解度については, すべての実験参加者が可視化表現の時系列で表現されていることを理解したという回答が得られた. しかし, 特徴語の文字の見づらさや, 特徴語とニュースの関連性が読みとれないことを指摘した. 「多くのニュースを時系列で見られる点では, ただ順番に見せられるよりはよく感じた」という意見も得られた. (3) インタフェースの使用感については, 可視化表現の拡大と縮小の操作性について困難であるという意見が 8 人中 5 人から得られた.

実験結果から, 実験参加者は, 3 章で想定したとおり, 提案するシステムを用いて時系列順に探索を行う傾向が確認された. 実験後の聞き取り調査での回答においても, 時系列で整理された情報が提示されていることを評価する発言があったことから, このシステムのデザインの妥当性が示唆された. 一方で, 特徴語のデザインと拡大・縮小機能の操作性については改善の余地があることが指摘されたことから, 今後改善を図る必要がある.

7. おわりに

本稿では, 長期間報道されるニュースの理解支援を目的として, 一連のニュース記事をニュース記事の報道時期とトピックの可視化表現と合わせて提示するインタフェースを提案した.

長期間報道される一連のニュース記事集合について, トピックごとに分類し, トピックの流れを特徴語により可視化した. しかし, 今回の実験で実施した聞き取り調査においては, トピック分類についてそれぞれのトピックが示すニュース内容の違いが不明であるという回答が得られた. 現在の実装では, 個々のトピック中のニュース内容の変化を読み取ることは可能であるものの, トピックを横断して探索する際にトピック間の関連性を把握することが困難である. そのため, トピックの分岐や収束をわかりやすくユーザに提示する必要があると考える. 今後, 図 2 に示したようなトピックの分岐及び収束を求める可視化表現の実現方法について検討していく.

謝辞

本研究では, 「毎日新聞 2016 データ集」の新聞記事コーパスを使用した. 記して謝意を表す.

参考文献

- [1] Card, S. K., Mackinlay, J. D. and Shneiderman, B.: *Readings in information visualization: using vision to think*, Morgan Kaufmann (1999).
- [2] Le, Q. and Mikolov, T.: Distributed representations of sentences and documents, *International conference on machine learning*, pp. 1188–1196 (2014).
- [3] Mikolov, T., Sutskever, I., Chen, K., Corrado, G. and Dean, J.: Distributed representations of words and phrases and their compositionality, *arXiv preprint*

- arXiv:1310.4546* (2013).
- [4] Shen, D., Wang, G., Wang, W., Min, M. R., Su, Q., Zhang, Y., Li, C., Henao, R. and Carin, L.: Baseline Needs More Love: On Simple Word-Embedding-Based Models and Associated Pooling Mechanisms, *arXiv preprint arXiv:1805.09843* (2018).
 - [5] Shneiderman, B.: The eyes have it: A task by data type taxonomy for information visualizations, *The craft of information visualization*, Elsevier, pp. 364–371 (2003).
 - [6] 小川恒平, 永塚光一, 渥美雅保: ニューラルネットワークを用いたニュース記事の抽出的三行要約, 人工知能学会第33回全国大会論文集, No. 4Rin111 (2019).
 - [7] 谷口祐太郎, 小林哲則, 林 良彦: 段階的なテーマ詳細化によるニュース情報の体系的獲得, 第18回インタラクティブ情報アクセスと可視化マイニング研究会予稿集, Vol. 18, No. 04, pp. 19–26 (2018).
 - [8] 松本直彰, 湯本高行, 山本岳洋, 大島裕明: 時間的推移を考慮した報道トピックの抽出と再構成, 情報処理学会研究報告, Vol. 2020-IFAT-139, No. 1, pp. 1–6 (2020).
 - [9] 鈴木聖乃, 岩崎有基, 松下光範: 人物関係の可視化におけるニュースストリームの理解支援, 第17回情報科学技術フォーラム講演論文集, No. J-032 (2018).
 - [10] 大島裕明, 中村聡史, 田中克己: SlothLib: Web 検索研究のためのプログラミングライブラリ, 日本データベース学会 letters, Vol. 6, No. 1, pp. 113–116 (2007).

競争力強化といった課題にいかに取り組み, 「アベノミクスの再加速」を実現できるかが問われている。2017年度予算編成に向けた各省庁の概算要求が31日, 出そろった。一般会計の要求総額は3年連続で100兆円を突破し, 101兆円台となる見通し。消費税率10%への引き上げを再延期する税制改正法案が18日, 衆院本会議で審議入りした。消費税率10%への引き上げ時期を2017年4月から19年10月に再延期する税制改正法案が8日, 衆院本会議で与党の賛成多数により可決, 衆院を通過した。2016年度の国の一般会計税収が, 法人税収の減少を主因に, 7年ぶりに前年度実績(56兆2854億円)を下回る見通しとなった。

付 録

A.1 提出されたレポートの一例(キーワード: 消費税+消費税増税)

軽減額は消費額が多い高所得世帯ほど増える一方, 負担感を緩和する効果は低所得者ほど大きくなる。低所得世帯ほど, 所得に占める消費税負担額の比率が大きいため, 軽減税率が負担率を押し下げる効果が高まる。軽減税率の主な対象は酒類と外食を除く飲食料品, 週2回以上発行している新聞の定期購読料で, 税率を8%に据え置く。食はテーブルなどの設備のある場所や, 顧客が指定した場所で調理や給仕を伴う飲食の提供と定義した。経団連の榊原定征会長は会見で, 来年4月に予定される消費税率10%への引き上げについて「社会保障の充実と財政健全化のためにも計画通り実施すべきだ」と述べた。首相は増税そのものは「重大な事態が発生しない限り確実に実施する」と強調。政府は23日発表した3月の月例経済報告で, 全体の景気判断を昨年10月以来, 5カ月ぶりに下方修正した。2月の「このところ一部に弱さもみられるが, 緩やかな回復基調が続いている」から「一部に」を削除した。首相は, 来年4月の消費税率10%への引き上げを再延期する考えを伝えた。延期期間は2019年10月までの2年半とすることを提案したが, 異論が出たため, 引き続き協議することとした。一方で「現在の消費低迷は社会保障制度の先行き懸念など将来不安が根本にあり, 増税を延期しても長期的な消費活性化にはつながらない」との声もあり, 評価と懸念が入り交じっている。国内景気の停滞感が強まる中, デフレ脱却と財政再建の両立や, エネルギー政策の立て直し, 農業の